2163

```
        RAW SEQUENCE LISTING                    DATE: 02/28/2002
        PATENT APPLICATION:  US/09/910,071      TIME: 09:28:42

        Input Set : A:\PTO.VSK.txt
        Output Set: N:\CRF3\02282002\I910071.raw

     3 <110> APPLICANT: Tomikawa, Mayumi
     4        Aikawa, Seiichi
     5        Matsuzawa, Fumiko
     7 <120> TITLE OF INVENTION: Method and Apparatus for Extracting and Evaluating Mutually
Similar
     8        Portions in One-Dimensional Sequences in Molecules and/or Three-Dimensional
     9        Structures of Molecules
    11 <130> FILE REFERENCE: 522.1921D2
    13 <140> CURRENT APPLICATION NUMBER: 09/910,071
    14 <141> CURRENT FILING DATE: 2001-07-23
    16 <150> PRIOR APPLICATION NUMBER: US 08/014,867
    17 <151> PRIOR FILING DATE: 1993-02-08
    19 <160> NUMBER OF SEQ ID NOS: 20
    21 <170> SOFTWARE: PatentIn version 3.1
    23 <210> SEQ ID NO: 1
    24 <211> LENGTH: 37
    25 <212> TYPE: PRT
    26 <213> ORGANISM: human
    28 <400> SEQUENCE: 1
    30 Gly Asp Val Glu Lys Gly Lys Lys Ile Phe Ile Met Lys Cys Ser Gln
    31 1               5                   10                  15
    34 His Thr Val Glu Gly Gly Lys His Lys Thr Gly Pro Asn Leu His Gly
    35             20                  25                  30
    38 Leu Phe Gly Arg Lys
    39          35
    42 <210> SEQ ID NO: 2
    43 <211> LENGTH: 39
    44 <212> TYPE: PRT
    45 <213> ORGANISM: bacterium
    47 <400> SEQUENCE: 2
    49 Glu Gly Asp Asp Ala Ala Ala Gly Glu Lys Val Ser Lys Lys Cys Leu
    50 1               5                   10                  15
    53 Ala Cys His Thr Phe Asp Gln Gly Gly Ala Asn Lys Val Gly Pro Asn
    54             20                  25                  30
    57 Pro Asn Leu Phe Gly Val Phe
    58          35
    61 <210> SEQ ID NO: 3
    62 <211> LENGTH: 51
    63 <212> TYPE: PRT
    64 <213> ORGANISM: rat
    66 <400> SEQUENCE: 3
    68 Met Ser Leu Ala Ile Leu Arg Val Ile Arg Leu Val Arg Val Phe Arg
    69 1               5                   10                  15
    72 Ile Phe Lys Leu Ser Arg His Ser Lys Gly Leu Gln Ile Leu Gly Arg
```

RAW SEQUENCE LISTING   DATE: 02/28/2002
PATENT APPLICATION: US/09/910,071   TIME: 09:28:42

            Input Set : A:\PTO.VSK.txt
            Output Set: N:\CRF3\02282002\I910071.raw

```
73              20                25                30
76 Thr Leu Lys Ala Ser Met Arg Glu Leu Gly Leu Leu Ile Phe Phe Ile
77        35                40                45
80 Gly Val Val
81     50
84 <210> SEQ ID NO: 4
85 <211> LENGTH: 142
86 <212> TYPE: PRT
87 <213> ORGANISM: unknown
89 <220> FEATURE:
90 <223> OTHER INFORMATION: Figure 23A
92 <400> SEQUENCE: 4
94 Thr Glu Glu Gln Ile Ala Glu Phe Lys Glu Ala Phe Ser Leu Phe Asp
95 1                5                10                15
98 Lys Asp Gly Asp Gly Thr Ile Thr Thr Lys Glu Leu Gly Thr Val Met
99        20                25                30
102 Arg Ser Leu Gly Gln Asn Pro Thr Glu Ala Glu Leu Gln Asp Met Ile
103        35                40                45
106 Asn Glu Val Asp Ala Asp Gly Asn Gly Thr Ile Asp Phe Pro Glu Phe
107     50                55                60
110 Leu Thr Met Met Ala Arg Lys Met Lys Asp Thr Ser Glu Glu Glu Ile
111 65                70                75                80
114 Arg Glu Ala Phe Arg Val Phe Asp Lys Asp Gly Asn Gly Tyr Ile Ser
115                85                90                95
118 Ala Ala Glu Leu Arg His Val Met Thr Asn Leu Gly Glu Lys Leu Thr
119              100                105                110
122 Asp Glu Glu Val Asp Glu Met Ile Arg Glu Ala Asn Ile Asp Gly Asp
123        115                120                125
126 Gly Gln Val Asn Tyr Glu Glu Phe Val Gln Met Met Thr Ala
127     130                135                140
130 <210> SEQ ID NO: 5
131 <211> LENGTH: 159
132 <212> TYPE: PRT
133 <213> ORGANISM: unknown
135 <220> FEATURE:
136 <223> OTHER INFORMATION: Figure 23B
138 <400> SEQUENCE: 5
140 Ala Met Asp Gln Gln Ala Glu Ala Arg Ala Phe Leu Ser Glu Glu Met
141 1                5                10                15
144 Ile Ala Glu Phe Lys Ala Ala Phe Asp Met Phe Asp Ala Asp Gly Gly
145              20                25                30
148 Gly Asp Ile Ser Thr Lys Glu Leu Gly Thr Val Met Arg Met Leu Gly
149        35                40                45
152 Gln Asn Pro Thr Lys Glu Glu Leu Asp Ala Ile Ile Glu Glu Val Asp
153     50                55                60
156 Glu Asp Gly Ser Gly Thr Ile Asp Phe Glu Glu Phe Leu Val Met Val
157 65                70                75                80
160 Arg Gln Met Lys Glu Asp Ala Lys Gly Lys Ser Glu Glu Glu Leu Ala
161              · 85                90                95
```

```
164 Asp Cys Phe Arg Ile Phe Asp Lys Asn Ala Asp Gly Phe Ile Asp Ile
165             100                 105                 110
168 Glu Glu Leu Gly Glu Ile Leu Arg Ala Thr Gly Glu His Val Thr Glu
169         115                 120                 125
172 Glu Asp Ile Glu Asp Leu Met Lys Asp Ser Asp Lys Asn Asn Asp Gly
173     130                 135                 140
176 Arg Ile Asp Phe Asp Glu Phe Leu Lys Met Met Glu Gly Val Gln
177 145                 150                 155
180 <210> SEQ ID NO: 6
181 <211> LENGTH: 28
182 <212> TYPE: PRT
183 <213> ORGANISM: unknown
185 <220> FEATURE:
186 <223> OTHER INFORMATION: calmodulin probe site 81-108, Figure 25, target
188 <400> SEQUENCE: 6
190 Leu Ala Asp Cys Phe Arg Ile Phe Asp Lys Asn Ala Asp Gly Phe Ile
191 1           5                 10                  15
194 Asp Ile Glu Glu Leu Gly Glu Ile Leu Arg Ala Thr
195             20                  25
198 <210> SEQ ID NO: 7
199 <211> LENGTH: 28
200 <212> TYPE: PRT
201 <213> ORGANISM: unknown
203 <220> FEATURE:
204 <223> OTHER INFORMATION: calmodulin probe site 81-108, probe, Figure 25
206 <400> SEQUENCE: 7
208 Ile Arg Glu Ala Phe Arg Val Phe Asp Lys Asp Gly Asn Gly Tyr Ile
209 1           5                 10                  15
212 Ser Ala Ala Glu Leu Arg His Val Met Thr Asn Leu
213             20                  25
216 <210> SEQ ID NO: 8
217 <211> LENGTH: 27
218 <212> TYPE: PRT
219 <213> ORGANISM: unknown
221 <220> FEATURE:
222 <223> OTHER INFORMATION: Calmodulin probe site 117-143, Figure 26, target
224 <400> SEQUENCE: 8
226 Ile Glu Asp Leu Met Lys Asp Ser Asp Lys Asn Asn Asp Gly Arg Ile
227 1           5                 10                  15
230 Asp Phe Asp Glu Phe Leu Lys Met Met Glu Gly
231             20                  25
234 <210> SEQ ID NO: 9
235 <211> LENGTH: 27
236 <212> TYPE: PRT
237 <213> ORGANISM: unkown
239 <400> SEQUENCE: 9
241 Val Asp Glu Met Ile Arg Glu Ala Asn Ile Asp Gly Asp Gly Gln Val
242 1           5                 10                  15
245 Asn Tyr Glu Glu Phe Val Gln Met Met Thr Ala
```

RAW SEQUENCE LISTING   DATE: 02/28/2002
PATENT APPLICATION: US/09/910,071   TIME: 09:28:42

        Input Set : A:\PTO.VSK.txt
        Output Set: N:\CRF3\02282002\I910071.raw

```
246          20                25
249 <210> SEQ ID NO: 10
250 <211> LENGTH: 8
251 <212> TYPE: PRT
252 <213> ORGANISM: unknown
254 <220> FEATURE:
255 <223> OTHER INFORMATION: Coordinates of C-alpha corresponding to the aminon acid
residue N
256       os. 7 to 14 in elongation factor of protein which is a binding si
257       te for phosphoric acid of GTP
259 <400> SEQUENCE: 10
261 Gly His Val Asp His Gly Lys Thr
262 1              5
265 <210> SEQ ID NO: 11
266 <211> LENGTH: 8
267 <212> TYPE: PRT
268 <213> ORGANISM: unknown
270 <220> FEATURE:
271 <223> OTHER INFORMATION: adenylate kinase - target
273 <400> SEQUENCE: 11
275 Gly Ala Pro Gly Ser Gly Lys Gly
276 1              5
279 <210> SEQ ID NO: 12
280 <211> LENGTH: 8
281 <212> TYPE: PRT
282 <213> ORGANISM: unknown
284 <220> FEATURE:
285 <223> OTHER INFORMATION: ras protein - target
287 <400> SEQUENCE: 12
289 Gly Ala Gly Gly Val Gly Lys Ser
290 1              5
293 <210> SEQ ID NO: 13
294 <211> LENGTH: 8
295 <212> TYPE: PRT
296 <213> ORGANISM: unknown
298 <220> FEATURE:
299 <223> OTHER INFORMATION: adenylate kinase (3ADK)
301 <400> SEQUENCE: 13
303 Gly Gly Pro Gly Ser Gly Lys Gly
304 1              5
307 <210> SEQ ID NO: 14
308 <211> LENGTH: 223
309 <212> TYPE: PRT
310 <213> ORGANISM: unknown
312 <220> FEATURE:
313 <223> OTHER INFORMATION: amino acid sequence of trypsin
315 <400> SEQUENCE: 14
317 Ile Val Gly Gly Tyr Thr Cys Cys Ala Asn Thr Val Pro Tyr Gln Val
318 1              5                    10                  15
321 Ser Leu Asn Ser Gly Tyr His Phe Cys Gly Gly Ser Leu Ile Asn Ser
```

```
322              20            25            30
325 Gln Trp Val Val Ser Ala Ala His Cys Tyr Lys Ser Gly Ile Gln Val
326         35            40            45
329 Arg Leu Gly Glu Asp Asn Ile Asn Val Val Glu Gly Asn Glu Gln Phe
330     50            55            60
333 Ile Ser Ala Ser Lys Ser Ile Val His Pro Ser Tyr Asn Ser Asn Thr
334 65            70            75            80
337 Leu Asn Asn Asp Ile Met Leu Ile Lys Leu Lys Ser Ala Ala Ser Leu
338             85            90            95
341 Asn Ser Arg Val Ala Ser Ile Ser Leu Pro Thr Ser Cys Ala Ser Ala
342             100           105           110
345 Gly Thr Gln Cys Leu Ile Ser Gly Trp Gly Asn Thr Lys Ser Ser Gly
346         115           120           125
349 Thr Ser Tyr Pro Asp Val Leu Lys Cys Leu Lys Ala Pro Ile Leu Ser
350     130           135           140
353 Asp Ser Ser Cys Lys Ser Ala Tyr Pro Gly Gln Ile Thr Ser Asn Met
354 145           150           155           160
357 Phe Cys Ala Gly Tyr Leu Glu Gly Gly Lys Asp Ser Cys Gln Gly Asp
358             165           170           175
361 Ser Gly Gly Pro Val Val Cys Ser Gly Lys Leu Gln Gly Ile Val Ser
362             180           185           190
365 Trp Gly Ser Gly Cys Ala Gln Lys Asn Lys Pro Gly Val Tyr Thr Lys
366         195           200           205
369 Val Cys Asn Tyr Val Ser Trp Ile Lys Gln Thr Ile Ala Ser Asn
370     210           215           220
373 <210> SEQ ID NO: 15
374 <211> LENGTH: 239
375 <212> TYPE: PRT
376 <213> ORGANISM: unknown
378 <220> FEATURE:
379 <223> OTHER INFORMATION: amino acid sequence of Elastase
381 <400> SEQUENCE: 15
383 Val Val Gly Gly Thr Glu Ala Gln Arg Asn Ser Trp Pro Ser Gln Ile
384 1           5             10            15
387 Ser Leu Gln Tyr Arg Ser Gly Ser Ser Trp Ala His Thr Cys Gly Gly
388         20            25            30
391 Thr Leu Ile Arg Gln Asn Trp Val Met Thr Ala Ala His Cys Val Asp
392         35            40            45
395 Arg Glu Leu Thr Phe Arg Val Val Val Gly Glu His Asn Leu Asn Gln
396     50            55            60
399 Asn Asn Gly Thr Glu Gln Tyr Val Gly Val Gln Lys Ile Val Val Pro
400 65            70            75            80
403 Tyr Trp Asn Thr Asp Asp Val Ala Ala Gly Tyr Asp Ile Ala Leu Leu
404             85            90            95
407 Arg Leu Ala Gln Ser Val Thr Leu Asn Ser Tyr Val Gln Leu Gly Val
408             100           105           110
411 Leu Pro Arg Ala Gly Thr Ile Leu Ala Asn Ser Pro Cys Tyr Ile Thr
412         115           120           125
415 Thr Gly Trp Gly Leu Thr Arg Thr Asn Gly Gln Leu Ala Gln Thr Leu
```

VERIFICATION SUMMARY   DATE: 02/28/2002
PATENT APPLICATION: US/09/910,071   TIME: 09:28:43

Input Set : A:\PTO.VSK.txt
Output Set: N:\CRF3\02282002\I910071.raw